network, the demand for that low bidder's product will necessarily increase. As demand increases, supply will be unable to increase at an equally rapid rate and prices will begin to rise. Given that the price quotations received by the engineering team reflect current supply and demand for large projects, the jump in demand will necessarily result in an increase in price.

AT&T has submitted the testimony of John C. Donovan in an attempt rebut existing criticisms of the low input values in HM 5.0.[33] His testimony includes a chart purporting to show that only two of 30 selective outside plant input values in HM 5.0 are the same as the lowest vendor quote. A careful review of Mr. Donovan's chart and the summary of Mr. Fassett's "validation data" set forth in Exhibit JCD-2 to his testimony undermines this effort.

First, Mr. Donovan does not address the original problem with HM 3.1. That version of the Hatfield Model consistently used the lowest available quote to set outside plant values. It remains unrebutted that, based on the vendor information considered by the engineering team at the time that HM 3.1 was released, they almost always used the lowest quote. No information obtained after the release of HM 3.1 can change that fact. It now appears that the outside plant engineering team obtained and considered additional quotes after the release of HM 3.1 and before the release of HM 5.0. However, none of the default values in the chart (attached hereto as Attachment 10) changed from HM 3.1 to HM 5.0. The lowest quotations (as of the release of HM 3.1) remained the basis for the default values in HM 5.0, despite Mr. Donovan's recent data gathering efforts.

Second, the chart sponsored by Mr. Donovan is yet another example of AT&T's misleading use of selective data. For instance, Mr. Donovan's chart ignores vendor estimates relating to some

[33] Rebuttal Testimony of John C. Donovan, Before the Alabama Public Service Commission, Docket No.25980, February 3, 1998.

significant elements of outside plant that are unfavorable to AT&T's position. Mr. Fassett received quotations for the cost of placing aerial drop lines and for estimates of the average length of drop lines. These two items are important cost drivers for the outside plant. The Inputs Portfolios for both HM 3.1 and HM 5.0 contain default input values for those items. Those input values are significantly lower that the quotes received in Mr. Fassett's survey. The quotes from Mr. Fassett's survey do not support the thesis of Mr. Donovan's testimony. Perhaps for that reason, Mr. Donovan conveniently neglected to include either "average drop length" or "aerial drop placement" in his validation chart. Mr. Fassett also received three quotes of approximately $11,000 from one vendor for the material cost of a manhole. Mr. Donovan did not put this unfavorable quote in his chart, either.

Much of the data provided by Mr. Donovan cannot be verified. For example, the chart indicates that the default values for an "aerial strand mounted block terminal" (item no. 7) and "buried pedestal block terminal" (item no. 8) are higher than the lowest quotes for those items. However, it is impossible to verify from the Model Inputs Portfolio that Mr. Donovan used the correct default value in his calculations. This is because the Inputs Portfolio does not contain separate values for the cost of only the terminals. The Inputs Portfolio combines the material and labor costs for these items, and does not explain how the two can be separated. The same kind of unverifiable data was used to tout the reasonableness of the default values for all trenching, item nos. 21-26 in the chart. Once again, we cannot test Mr. Donovan's calculations because it is impossible to discern from the Inputs Portfolio what were the default values for trenching.

In other cases, the default values in Mr. Donovan's chart are plainly wrong. The accurate default values for item nos. 19 and 20, which relate to the cost of manhole excavation and backfill in suburban and metropolitan areas, are cost ranges. The range in the suburban environment was $3,200-$3,500; in the metro environment, it was $3,500-$5,000. Mr. Donovan, however, chose to show only the higher number in his chart as the default value -- skewing his results. Other times, where a vendor gives more than one quote for more than one state, Mr. Donovan uses the lowest quote and ignores the higher one. By doing this, Mr. Donovan was able to increase, in a misleading way, the percentage by which the default value for these items purportedly was "higher than the lowest" quote.

Similarly, the default values relating the pole labor and total pole investment are wrong. Mr. Donovan indicates that the default value in HM 5.0 for pole labor is $216, which was based on the quote submitted by vendor "h". What Mr. Donovan fails to acknowledge is that vendor "h" did not submit a quote of $216 for pole "labor", as defined by AT&T. Even though none of the vendors that submitted quotes for pole labor stated that the costs of anchors and guys were included in their quotes, AT&T has taken the position that the labor cost of a pole should include the cost of material and labor for "exempt materials" such as pole anchors and guys. Vendor "h" submitted a separate quote of $231 for material and labor costs associated with anchor and guys. Thus, the quote from vendor "h" for pole "labor", as defined by AT&T, was actually $447.

Still other data was improperly grouped by Mr. Donovan when he made his calculations. Turning again to the quotes relating to trenching -- item nos. 21-26 -- Mr. Donovan indicated that Mr. Fassett requested and obtained quotes for the cost of digging trenches that were both six inches and 12 inches wide. Not surprisingly, the cost of digging a six inch wide trench proved to be less expensive than digging one 12 inches wide. These items were separately quoted by the vendors, and should have been the subject of separate calculations by Mr. Donovan. Instead, he treated all quotes for trenching the same, again skewing the results in his favor. In another case, Mr. Donovan ignored price quotes for the cost of plowing cable having a diameter of 1.51-2.44 inches, even though Mr. Fassett specifically asked for and received quotes for plowing this kind of cable. Mr. Donovan only listed lower price quotes relating to cable a smaller diameter. Using the omitted data should have resulted in a higher default value for plowing cable, and would have changed Mr. Donovan's calculations.

C. The Model Adopts Unsupported Input Values

The Hatfield Model obtains enormous cost savings through presumed efficiency gains that will supposedly become possible in a more competitive environment. The efficiency gains often come in the form of undocumented and unverifiable assumptions regarding the competitive environment and are only supported by the Hatfield modelers' "expert opinions." Several examples follow:

1. Structure Sharing

Significant cost savings are made possible by the Model's structure sharing assumptions, whereby savings might arise from better cooperation with other utilities. The Model's sharing assumptions bear no resemblance to the actual structure sharing practices of GTE, are not technically feasible, and distort economic incentives. To achieve such dramatic structure sharing percentages, the Hatfield Model assumes not only a "scorched-node" approach to the telephone network, but essentially also a scorching of the existing power and cable networks. Without "total utility scorching," it would be impossible for ILECs to share the cost of placing the new futuristic Hatfield network. The Hatfield Model assumes that in all but the two lowest density zones, the ILEC will bear only 25% of the costs of aerial structure. The Hatfield modelers' assumptions purportedly are based upon a statement in a New York Public Service Commission proceeding that the sharing of poles among six entities "would not be uncommon." But there was no suggestion in this document that such a practice would be common in New York, let alone in Alabama.[34] Significantly, the Hatfield modelers fail to consider the additional costs associated with such sharing. On the very same page cited by the Hatfield modelers, the utility panel goes on to state that "a forty foot pole may require rearrangement."[35]

The Hatfield Model estimates that the ILEC will assume only 33% of the costs of buried cable. According to Hatfield, "no charge developer-dug trenches reduces greatly the effective portion of total buried structure borne by the LEC." This reliance upon developer-dug trenches is misguided. For such costs savings to be realized, the utility and CATV providers must also "scorch" their facilities. Moreover, HM 5.0 does not consider engineering requirements associated with sharing buried applications. For example, the AT&T Outside Plant Engineering Handbook specifies that *"joint trenching with power facilities should be employed only for distribution cables and service wires, not for feeder or trunk cables."*[36] HM 5.0, however, assumes that ILECs will share trenches in the feeder portion of the network with power utilities. AT&T's Outside Plant Handbook also specifies the minimum separation distances that engineers must maintain between power and

[34] All that was said is that "a typical 40 foot distribution pole can generally accommodate up to 2 third party communications attachments and 2 third party co-lashes." (Direct Testimony of Electric Utility Panel, Public Service Comm. Of the State of New York, Case No. 95-C-0341 at 14 (attached hereto as Attachment 11.) There is no differentiation between what is possible and what is probable. Moreover, in a recent draft decision, the California Public Utilities Commission ruled that the structure sharing percentages included in the Hatfield Model are not reasonable. Draft Decision ALJ McKenzie, Before the Public Utilities Commission of the State of California, Docket Nos. R.93-04-003 and I.93-04-002, Mailed December 23, 1997, Page 35 (attached hereto as Attachment 12).

[35] Direct Testimony of Electric Utility Panel, Public Service Comm. Of the State of New York, Case No. 95-C-0341 at 14, Attachment 11.

[36] AT&T Outside Plant Engineering Handbook, August 1994, Sec. 9, p. 6, Attachment 4.

telephone plant crossings to ensure public safety and the integrity of the facilities. The separation distance is twelve inches in most cases. More importantly, the buried PIC cable recommendations in AT&T's and some ILECs' Outside Plant Engineering Manuals indicate that power lines must be placed at a greater depth than the telephone plant, which substantially increases the cost of trenching. Again, HM 5.0 ignores these established engineering standards, and in so doing, understates ILEC costs.

For most density zones, the Hatfield Model assumes that GTE will bear only 33% of the costs of underground conduit. This assumption is based upon a study of New York City subway ducts, which are occupied by "well over 300 telecommunication providers." Upon further analysis, however, it becomes clear that the ILEC's cables comprise 20,073 sheath miles in New York State (excluding 7,154 conduit miles in Manhattan/Bronx) while all other users comprise only 151 miles. There may be numerous "telecommunication providers" in New York, but they hardly share 67% of the conduit, as HM 5.0 assumes.

2. Fill Factors

The Hatfield Model is static in that it fails to recognize that technological progress under competition will have important consequences for the rate at which network facilities are utilized. The problem of optimally investing in discrete plant when there is growth has a component not found in static situations. In his 1978 paper in the *Review of Economic Studies*, David Starrett shows that the cost-minimizing firm in a dynamic situation trades off some spare capacity against the economies of scale in construction. The firm minimizes cost by choosing the lengths of the intervals between which it invests. During periods between investments there will often be significant spare capacity, so it is often optimal and cost minimizing to have substantial spare capacity.

In this respect, the Model assumes fill factors (i.e., utilization rates) that are too high. The Hatfield Model's use of these high fill factors causes costs to be understated because the fill factor, in part, determines how much cable is needed. HM 5.0 appears to be based on the belief that competitive firms should have minimal spare capacity. In fact, the opposite is true; in a competitive environment, GTE may be required to have more, not less, spare capacity. The FCC's findings on spare capacity in interstate long-distance demonstrate this point.

> . . . MCI and Sprint alone can absorb overnight as much as fifteen percent of AT&T's total 1993 switched demand at no incremental capacity cost; within 90 days MCI, Sprint, LDDS/Wiltel, using their existing equipment, could absorb almost one-third of AT&T's total switched capacity; or that within twelve months, AT&T's largest competitors could absorb almost two thirds of total switched traffic for a combined investment of $660 million. We therefore conclude that AT&T's competitors have sufficient excess capacity available to constrain AT&T's pricing behavior.[37]

To cast the FCC findings in terms relevant to the current discussion, note that MCI and Sprint combined are roughly one-half of AT&T's size. Overnight they can each absorb about 15 percent of AT&T's capacity. This implies that MCI and Sprint together have at least 30 percent spare capacity that could be deployed overnight. The implication of these findings is that the uncertainty that is accompanied by the introduction of competition may result in more rather than less spare capacity if it requires firms to be flexible enough to respond to the vicissitudes of the market.

There is no single *optimal* fill factor. The *optimal* fill factor will depend on growth rates, interest rates, modularity, uncertainty, and depreciation. Properly modeled fill factors increase over time, as demand increases to fill capacity. Because of its static nature, the Hatfield Model cannot properly determine fill factors.

---

[37] Federal Communications Commission, In the Matter of Motion of AT&T Corp. to be reclassified a s a Non-Dominant Carrier, FCC 95-427, October 15, 1995, paragraph 59.

The fill factors employed by the ILECs serve as an excellent starting point. Not only have they developed over time, but also were constantly overseen by state and federal regulatory bodies to guarantee appropriate times to install new services in the face of growing but uncertain demand. They were also developed to meet important service quality requirements.

Finally, HM 5.0 sizes its fiber feeder cables using a fill factor of 100%. The result is a fiber network with no excess capacity. No interoffice or outside plant engineer would construct a fiber plant in this manner because the ILEC's network would be unable to handle any short-term demand fluctuations. In addition, there must be maintenance spares available for situations where copper cable pairs or fiber strands become inoperable and require replacement. HM 5.0 simply ignores these standard-engineering practices.

3. Structure Mix

The structure mix assumed in HM 5.0 is based solely on "expert opinion which have not been substantiated by any analysis or empirical study.[38] With the release of Version 3.1, the Hatfield modelers significantly increased the amount of aerial placement in urban areas to 85%. This was despite the fact that AT&T's Outside Plant Handbook instructs that aerial plant should be used only as a last resort, in cases in which buried and underground plant is not feasible. This figure is still being used by the latest release of the Model. The developers' justification for this assumption is that "block cable," which runs to high rise buildings, is included as a subset of an ILEC's aerial account. Assuming that this is accurate, and that block cables are in fact considered "aerial," this fact thoroughly undercuts the corresponding HM 5.0 assumptions regarding placement costs. Block

[38] Hatfield Model, Release 5.0, Inputs Portfolio, HAI Consulting Inc., Boulder, Colorado, December 11, 1998, page 36

cable runs from the outside wall of a building, *under the sidewalks and streets*, and back up to the neighboring building. As such, block cable conduit placement cost should be no different -- and perhaps higher -- than underground conduit placement cost, which is assumed to be $75 per foot. By treating the block cable in the aerial category, the Hatfield supporters significantly reduce the Model's cost estimates.

In addition, HM 5.0 has added a new and inappropriate dynamic to the structure mix assumptions relative to outside plant. The new dynamic is called the "Buried Fraction Available for Shift". While the modelers suggest that this function is designed to allow economic choices between aerial and buried plant, especially in view of difficult soil conditions, it is actually used to shift plant in one direction only, away from buried and into less expensive aerial. This is inappropriate and inconsistent with forward looking trends. The Hatfield modelers correctly state, "Buried cable is now used wherever feasible..."[39] They also correctly state, "...the public clearly desires more out-of-site plant for both aesthetic and safety reasons."[40] Yet, with no support whatsoever, they have chosen to place 75% of the default values for buried plant "available for shift" in all but the highest two density zones, a completely arbitrary and unsupported assumption.

As a threshold matter, it is clear (and the Hatfield modelers acknowledge) that the forward looking trend is toward significantly more out-of-sight plant. In addition to the reasons cited above, buried and underground plant provides a much higher level of reliability than aerial plant because

---

[39] Hatfield Inputs Portfolio, p. 28

[40] Id.

hurricanes, ice storms, etc. do not have anywhere near the potential to disrupt service when plant is in the ground versus when it is in the air.

It appears that the addition of this "Buried Fraction Available for Shift" component is nothing more than a veiled attempt on the part of HM 5.0 developers to shift plant mix out of the higher cost (buried) category and into the lower cost (aerial) category. Model developers would be particulary inclined to design the Model this way given the recent recommendation of the FCC that the cost of cable buried with a plow be assigned 100% to the telephone companies.[41] Consequently, by shifting plant away from buried and into aerial, where structure sharing percentages recommended by the FCC are lower,[42] this feature promises to produce the Model developers desired result lower costs Model developers are seeking.

4. Expense Factors

The Model estimates the annual expenses for unbundled network elements by applying an ARMIS expense-to-investment factor. This factor divides current ARMIS expenses by their investment counterparts, and then applies these ratios to the Model's estimated investments. The problem with this method is that it assumes that if an investment is reduced, the expenses necessary to maintain that investment will be reduced in the same proportion, a relationship that often does not apply. Simply because a party pays less for a particular asset does not mean that its related expense will subsequently be decreased proportionately. For instance, the investment might reflect a discount on hardware, but not the subsequent maintenance contract. A vendor will frequently back-

[41] Further Notice of Proposed Rulemaking, Before the Federal Communications Commission, CC Docket Nos. 96-45 and 97-160, July 18, 1997, Paragraph 80.

[42] Id., Paragraph 81.

load a sale by discounting the initial purchase, but not the upgrades. The investment could reflect the price of relatively poor quality equipment, which will generally lead to higher future maintenance expenses. Indeed, there is often an *inverse* relationship between the price of an asset and the costs necessary to maintain it.

For certain USOA accounts, the modelers chose to override the ARMIS expense ratios using data derived from other sources and replace it with expert opinions or other studies. For example, it overrides the network operations expense ratio by assuming a 50% reduction in all current accounts in this category, including power expenses. The FCC has specifically noted Hatfield's failure to provide justification for its expense factor assumptions.[43] Furthermore, it overrides the switch maintenance expense by using a value derived from a New Hampshire study. This mix and match approach violates fundamental cost modeling principles.

An ALJ for the California Public Utilities Commission has agreed that Hatfield's use of the switch maintenance factor is inappropriate: In his draft decision rejecting the Hatfield Model in favor of the Pacific Bell cost model, an ALJ for the California Public Utilities Commission specifically criticized Hatfield Model's reliance upon New Hampshire data to determine switch maintenance expenses in California:

> we think that a maintenance factor derived from investment - which is almost certain to be less precise than a maintenance estimate based on actual experience - should be based upon data for a state with demographic and topographic characteristics reasonably comparable to California's. New Hampshire's are clearly not.[44]

[43] FCC Further Notice of Proposed Rulemaking, Docket 96-45 and 97-61160, July 15, 1997 ("FNPR") at 165.

[44] Draft Decision of ALJ McKenzie, Before the Public Utilities Commission of California, R.93-04-003 and I.93-04-002, Mailed 12/23/97, Page 32, Attachment 12.

Finally, the Model's method of calculating expenses is an example of what is called, in statistical or econometric parlance, "causal forecasting."[45] For example, when the Hatfield Model calculates an expense factor for a wire center building by dividing 1995 ARMIS-reported expenses associated with buildings by 1995 ARMIS reported investments in wire center buildings, it is estimating a *single* parameter of a *single* equation regression model with a *single* explanatory variable. In this example for building expenses, building investment is the sole explanatory variable.[46] It is essentially a specification of a "simple regression term;" a regression that does not include an intercept term and any other variable with explanatory power such as the number of switched access lines.

This approach is inadequate for a variety of reasons. The value of the regression coefficient is estimated as a ratio of ARMIS expense to ARMIS investment or to ARMIS reported lines. Thus, the expense part of the Model is calibrated or estimated using only *one* observation. In addition, the single variable regression specification is an assumption, one that needs to be tested statistically. Moreover, the factor approach produces a great deal of uncertainty in expense estimation.[47] Finally, and perhaps most important, while the factors are assumed constant for an ILEC in the Model, they should at least vary with the size of the ILEC.[48]

---

[45] That is, it assumes that whatever investments are decreased X%, the corresponding expenses will decrease in proportion - exactly X%.

[46] The mathematical equation is of the form $E = aI$, where E = expense, I = investment, and a = expense factor.

[47] This problematic aspect of the Model was identified during the recent depositions of Dr. Mercer and Ms. Murray, before the California Public Utility commission, Depositions of Robert Mercer, Joseph P. Riolo and Terry Murray, Docket Nos. R.93-04-003, I.93-04-002, March 7-8, 1997, p. 64.

[48] *Id.* at 139.

5. Common Costs

The Hatfield Model treats common costs by simply increasing all of its cost estimates by 10.4%. This treatment of common costs as an across-the-board increase in attributable cost is both arbitrary and incorrect. The argument provided by AT&T is a circuitous and meaningless justification for calculating GTE's variable, indirect overhead costs. Nowhere does it try to explain why the 10.4 % factor supposedly true of AT&T's own costs should also be true of GTE's costs. It is unclear why a LEC's common cost structure should correspond to a reasonable degree of approximation to the cost structure of a pure inter-exchange carrier. Indeed, the fact that the ILECs' networks have acknowledged scope economies in the provision of their myriad products and services strongly suggests higher common costs than exhibited by less capital-intensive, more specialized firms, such as AT&T was in 1994. Furthermore, if AT&T's access charges decrease, as is likely under the FCC Order, its ratio of common costs to total cost would *increase* well beyond 10.4%.

6. Network Operations Factor

According to Appendix D of the Model's documentation, there are six accounts of expenses that comprise Network Operating Expenses ("NOEs"). Network Administration (6532), Plant Testing (6533), and Plant Operations Administration (6534) expenses are the three largest accounts and are similar in magnitude. Engineering expenses (6535) are roughly half of those accounts. Power expenses (6531) are slightly lower than Engineering expenses (6535), and Provisioning expenses (6512) are the lowest by a wide margin.

The Model reduces these expense accounts by 50%, claiming that in a forward-looking environment, such reduction is warranted. However, Appendix D and the white paper written by Paul Hansen, discussed above, provide quantification supporting this reduction for only three of the six accounts -- Plant Testing, Plant Administration, and Engineering expense.[49] In his paper, the author estimates a 20% reduction in Plant Testing and Plant Administration expenses. For the third account, Engineering, he cites AT&T's outside plant expert, Ernest Carter, to support the notion that there would be at least a 25% reduction in regular Engineering expenses as a result of mechanization of record keeping. Even if these reduction estimates *were* valid they could hardly support the 50% across-the-board reduction factor in HM 5.0.

For the remaining three accounts (Network Administration, Power, and Provisioning), Hansen's support for the 50% reduction is either 1) theoretical technological advances[50], or 2) an unsubstantiated double counting argument. The technological advance argument, however, provides little support for the assertion that the Network Operations Factor (NOE) falls, and certainly none that it should fall by 50%. Indeed, there is evidence to the contrary. One need only look to estimates of total factor productivity (TFP) presented in the FCC price-cap review dockets. TFP is a reflection of the reduction in costs as a result of technological change. In these proceedings, TFP estimates ranged from 6% on the high side to less than 1% on the low side. Even if the total reduction in costs is roughly 6% per year, it would be very unlikely for all of the components that make up the NOE cost to fall at double digit rates.[51]

On the double counting point, the cost that Mr. Hansen thinks will be avoided are actually recovered in non-recurring charges. Deducting costs that are recovered through non-recurring charges from NOE does not accurately measure the costs of the ILECs' network. It confuses the

---

[49] White Paper on Network Operations Expense Factor (.50), Attachment 1.

[50] *Id.*

[51] Although , network operations forms only a part of the total costs considered in a total factor productivity, reference to TFP studies are valid since components of the costs tend to move together.

measurement and attribution of cost with cost recovery. Recovery of non-recurring costs is an entirely separate issue and confusing the two will lead to serious errors.

In the final analysis, the Hansen white paper provides strong support for the fact that the NOE will not be reduced in any appreciable fashion. The white paper shows a significant downward trend in RBOCs' NOEs, with adjustments for inflation, for the time period 1988 to 1996. In the eight year study period, RBOCs experienced a 46% reduction in NOE. This translates into a 9.3% *annual reduction* in NOE[52] -- strongly indicating that most of cost benefits attributable to technological advances are already reflected in the ARMIS report for recent years.

If NOEs have already fallen by approximately 50% in the last seven or eight years, to apply another 50% reduction would considerably exaggerate the effect of any technological change. Clearly the efficiency gains in network operations through recent technological advances by the ILECs are *already* reflected in the expenses reported in ARMIS. Mr. Hansen offers no compelling reason to assume that this downward trend will continue at the same rate. It simply is not rational to insist that there be *another* 50% reduction in NOEs in the future.

---

[52] [1-exp(In(0.46)/8)]

II. HM 5.0'S CUSTOMER LOCATION APPROACH IS FLAWED AND CLOSED IN NATURE.

The Model's developers claim that one of the major changes between HM 5.0 and its predecessors is a fundamental revision of the customer location methodology. According to the Model's documentation, the HM 5.0 input data locate customers much more precisely. These data determine the actual precise locations of as many customers as possible through latitude and longitude geo-coding of their addresses. [53] Furthermore, the documentation claims that because HM 5.0's approach identifies the actual locations (accurate to within 50 feet) of most telephone customers, it produces the most sophisticated demographic data set of its type. [54]

Our analysis of HM 5.0's revised customer location approach *could not* confirm these claims and revealed a series of fundamental flaws. HM 5.0's new customer location approach does *not* provide significant improvements over the alternative approaches taken in earlier versions of the Model. As pointed out on previous occasions, geo-coding on a national scale is highly imprecise and simply cannot be done with sufficient accuracy to provide a basis for calculating the costs of unbundled network elements or the size of the universal service subsidy.[55] Moreover, it appears that by clustering the geo-coded or otherwise assigned locations of customers, the Model developers essentially reverse the results of their geo-coding efforts in the end, and use a hypothetical customer distribution approach instead. In particular, regardless of the precision with which customer

[53] Hatfield Model, Release 5.0, Model Documentation, HAI Consulting Inc, Boulder, Colorado, December 11, 1997, page 5.

[54] *Id.*, at 23.

[55] Comments of GTE Corporations, Further Notice of Proposed Rule Making, CC-Docket 96-45, 97-160, September 2, 1997, pages 11.

locations are identified, the model itself places distribution plant under the assumption that locations are evenly distributed within distribution areas. This assumption is especially imprecise in low density areas. There is general agreement that customer locations tend to be along roads and roads in low density areas are often not evenly distributed.

A. Description Of HM 5.0's Customer Location Approach

As described in PNR's *Example of Customer Location: Raw Address Files to Clustered Output* and the *Model Description*, pages 21 through 29, the developers of the Hatfield Input Database go through a series of steps to determine the distribution architecture for each census block (CB). The following is a brief description of the development of the Hatfield input database. Due to the closed nature of the database, the following information is based on the documentation only and has not been validated by the authors of this report.

The process purportedly commences with Metromail Inc.'s National Consumer Database and Dun & Bradstreet's National Database for residential and business customer location counts, respectively. Centrus Desktop, a commercially available geo-coding software application, then compares the customer's street address as it appears in the input file to the address records contained in the USPS ZIP+4 directory and Geographic Data Technology's (GDT) enhanced street network files. Three scenarios can result from this process: Either the address is matched to United States Postal Service (USPS) files or the address is matched to USPS files *and* the GDT street network, or the address is not matched at all.

Under the first scenario, the ZIP+4 for the customer address is returned. The location information, however, is later discarded in the Model's customer location process and a surrogate method (described below) is used instead. Under the second scenario, Centrus Desktop determines a latitude and longitude for the customer's location to the X $^{th.}$ decimal place with an accompanying Census Block designation. For the purposes of the Hatfield Model, it is claimed that only geo-codes assigned at the 6th decimal place are used in determining customer locations. All other location information is dropped and the surrogate method is used instead. In the last scenario the surrogate method is always necessary.

Next, the target number of residential locations is determined by first eliminating duplicate records and then comparing total residential counts between the Claritas (an alternative database on US demographics) and Metromail databases.

The target number of business locations is supposedly determined by the Dun & Bradstreet National Database, and by simply adding 1 million surrogate points that are "believed to be missing."[56] Surrogate points consist of unlocated customers who are assumed to be located uniformly along the periphery of the Census Block. The "pseudo" geo-codes implied by these placements are subsequently added to the customer location file.

Once all estimated residential and business customers are either geo-coded or assigned to a surrogate point, a clustering algorithm essentially reverses the geo-coding efforts and aggregates all customers into a set of clusters.

Finally, yet another undefined algorithm, PointCode, is employed that supposedly translates between coordinate systems, computes distances and assigns additional characteristics to cluster records.

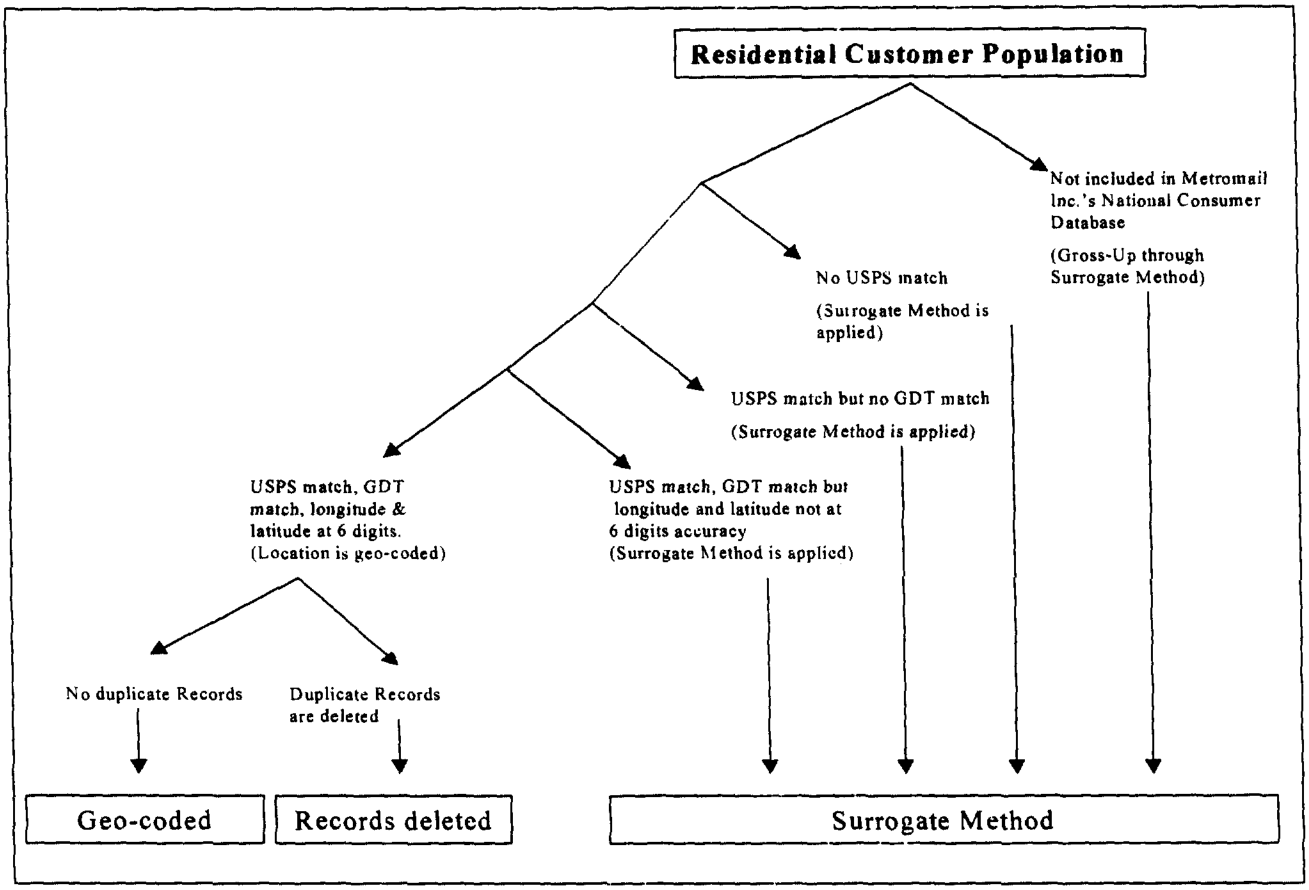
This process is illustrated in the figure below.

---

[56] "Example of Customer Location: Raw Address Files to Clustered Outputs" by PNR and Associates, date unknown, page 5.

Figure 1

HM 5.0 Customer Location Process



Only the final product of this process is subsequently included in the Model's input database and is therefore used in determining the TELRICs for various ILECs. We note that it is only at this point that an actual analysis of the Model can be conducted. All preceding steps are claimed to be either intellectual property, proprietary, or confidential.

B. Flaws In HM's Customer Location Approach

The success of geo-coding depends on three broad factors: 1) the integrity of the address match list, 2) the geo-coding engine, and 3) the accuracy and currency of the source street file. The following is a brief description of the major flaws that appear in the Model's customer location approach.

1. HM 5.0's Input Database Is Entirely Closed.

Paragraph 250 of the Report and Order of the Federal-State Joint Board on Universal Service states that the cost study or model and all underlying data, formulae, computations, and software associated with the model must be available to all interested parties for review and comment. All underlying data should be verifiable, engineering assumptions reasonable, and outputs plausible. The sponsors of the Hatfield Model have repeatedly claimed that all aspects of the Model are publicly available and open for inspection by third parties.

However, HM 5.0's Input Database clearly fails this criterion. First, all databases used in this process are considered intellectual property. By PNR's own estimates, it would cost a third party over $2.5 million in licensing agreements to review all of the databases that went into the customer location approach.[57] This cost figure does not include all external models and algorithms that were used in the process of determining the clusters. The sponsors and developers of the Model

[57] "PNR Estimates of the Resources Required to Support the Customer Location Model," PNR and Associates,. page 2.

even claim that certain intermediate results that lead to the final database are confidential and are thus unobtainable by third parties.[58]

Moreover, as PNR freely admits, it may take a new third-party processor 6-12 months to become fluent with the models and produce the first deliverables and the third party service bureau may not have the requisite understanding of the component data sources and their limitations to answer technical inquires or enhance the model. [59]

Given the closed nature of the Model's input database, it is not possible to conduct a thorough validation study of either the Model's input database or crucial cost drivers such as the length of the feeder and distribution cable, the density of clusters, etc.

2. The Vast Majority Of The Hatfield Model Is Pre-processed.

Our analysis of the new Model's input database reveals not only the Model's closed nature, but also its sheer size and complexity. While we did not have the opportunity to review all of the databases used in the customer location process, we gather from various documents that the input database must be a product of *at least* 12 different databases and 5 independent models or algorithms.[60] The major inputs to the Model are the result of massive pre-processing that can be neither analyzed or altered in a simple fashion. Thus, the actual Model becomes only the tip of the iceberg, and neither a review nor enhancements can be made to the Model's database.

---

[58] Affidavit of Richard N. Clarke, Before the Public Utilities Commission of the State of Minnesota, PUC Docket Nos., P-999/M-87-909, February 4, 1998, page 6.

[59] *Id.*

[60] We understand that PNR's clustering program was recently submitted to the FCC. The complexity of the process and the concomitant difficulty in performing independent evaluations is clearly illustrated by the fact that the program consists of 95 pages of code, programmed in C.

3. **Only a Portion of Customer Locations Are Actually Geo-coded.**

The Hatfield Model's description of geo-coding is misleading in that suggests that the *actual* locations of 95% of customers (accurate within 50 feet) were used in the Model. For instance, the developers claim that in general, geo-coding to the actual point location (i.e., sixth decimal place) is successful 70%-80% of the time.[61] What the documentation fails to point out is what the definition of in general means, which in this context is very important.

A close look at the Model's customer location approach, however, reveals that only a portion of actual customer locations are actually geo-coded and that none of this information is used to determine the TELRICs.

First, conflicting information exists on the actual address count contained Metromail Inc.'s National Consumer Database. As of December 5, 1997, Metromail Inc. reported that its database contained 74.4 million named and unnamed address records for the 50 states.[62] Contrasting this figure to the 1996 Bureau of the Census data of 109.8 million households shows that only 67.8% of households are actually being *considered* for geo-coding. On December 23, 1997, Metromail changed this statement and reported that its database contained not 74.4 million but 98.2 million address records.[63] This would imply that only 89.4% of households are *considered* for geo-coding. In Metromail's marketing brochure for the National Consumer Database, the company claims that

---

[61] "Example of Customer Location: Raw Address Files to Clustered Outputs" by PNR and Associates, Inc. page 4.

[62] *Ex Parte* Presentation - Proxy Cost Models, by Bell South, Sprint and U.S. West, CC-Docket No. 96-45, December 11, 1997, page 2. and January 9, 1998.

[63] RE: *Ex Parte* Presentation - Proxy Cost Models, by AT&T, CC-Docket No. 96-45, December 23, 1997, page 3.

the database consists of 103 million people, i.e., 95% of all U.S. households. The address count by state for the first two responses is provided in Appendix B.

It is unclear how many records Metromail's database actually contains.[64] What is clear however, is that the address list that is first even *considered* for geo-coding is incomplete. Moreover, the PNR documentation on geo-coding states that the Metromail database includes duplicate records. Thus, the actual count is believed to be even lower.

Second, not all addresses can be successfully geo-coded. Regardless of the level of accuracy, on average, there is only a 60% match rate (i.e., 6 out of 10 customer locations are successfully geo-coded). Generally, GDT - enhanced data can have match rates of up to 97% in highly urban areas[65]. In rural areas, however, these figures drop to roughly 50%. This is mainly due to the fact that some rural regions have not yet developed an E911 system. Rural areas also have a much lower hit rate because of the predominance of rural routes and post office boxes on such lists. The low hit rate in rural areas is particularly problematic for USF purposes, where the goal is to identify high cost areas.

Third, as stated in the Model documentation, HM 5.0 is only using location information that can be geo-coded to the 6th digit. This further reduces the total number of locations that are actually geo-coded. Based on PNR's own examples, it appears that roughly between 60% to 80% of all geo-codable locations can be geo-coded to the 6th digit.[66]

---

[64] The authors of this paper have contacted Metromail Inc. directly to obtain a quote on the number of addresses contained in the database but did not get a response in time to be included in this paper.

[65] These figures are based on the authors best knowledge and experience with geo-coding. We do not claim that this is the actual success rate that was achieved in the geo-coding efforts by PNR and Associates, Inc. We have asked for these figures from the Hatfield sponsors in a discovery request but have not received a response in time to be included in this paper.

[66] Example of Customer Location: Raw Address Files to Clustered Outputs" by PNR and Associates, date unknown, page 4.

Based on the information above, and assuming for a moment that PNR achieved a 100% success rate in "geo-codability," the Model actually geo-codes a range of merely 45% - 76% of all customer locations.[67] Including our concerns about the "geo-codability "of addresses, this range could drop to a low of roughly 20%.[68] In a recent affidavit filed by

Furthermore, the Model sponsors claim that their geo-coding exercise is accurate within 50 feet. [69] This statement is completely unsupported, and we seriously doubt its validity. Even the most sophisticated geo-coding software use address ranges and then makes judgments about the actual customer location. Addresses may be mapped onto the right road, but in rural areas will be no closer than within about 160 feet of actual customer location.[70]

4. Fundamental and Unchangeable Engineering Assumptions Are Used in Preprocessing the Data.

Despite the fact that the FCC's 9th criterion requires that critical engineering assumptions be open to examination (see Appendix C), the design of distributions areas, which is heavily based on engineering assumptions, takes place in the preprocessing of the data. Therefore, the critical assumptions of the maximum lines in a distribution area, maximum loop lengths, and maximum separation between customers are entirely beyond the reach of the Model's user. Tellingly, the Model's sponsors themselves admit that these engineering assumptions both are fundamental and cannot be changed when they explain that the maximum number of lines depends on DLC technology and that the clusters themselves might be subject to change if the technology changes.[71]

---

[67] Low: 0.75*1.0*0.6 = 45%, High: 0.95*1.0*0.8=76%.
[68] Low: 0.75*0.5*0.6 = 22.5%.
[69] Hatfield Model, Release 5.0, HAI Consulting Inc., Boulder, Colorado, December 11, 1997, page 23.
[70] Written statement by Etak Incorporation, January 20, 1998.
[71] *Model Documentation*, p. 27, footnote 30.

Similarly, the choice of a maximum of two miles between customers is an entirely unsupported assumption.[72]

5. HM 5.0's Input Database Does Not Make Use Of The Geo-coded Data.

While the Model's documentation leads the reader to believe that actual customer locations are being used to model telephone loops, in actuality the clustering algorithm along with the surrogate method essentially reverses these efforts and turns out customer distribution not much different from that in previous versions of the Hatfield Model. The fact that a significant number of customer locations are not geo-coded at all and are assigned to surrogate points makes this geo-coding exercise even more trivial. That is, customers are spaced evenly over the rectangular areas encompassed by the clusters. In large, low-density distribution areas, customers are therefore likely to be distributed into areas that are actually unserved, i.e., areas that contain no roads on which to locate dwelling units and business establishments.

6. The HM 5.0 Clusters Can Be Imprecise and May Violate The Engineering Rules That Are Supposed to Constrain The Establishment of Distribution Areas.

Complete access to the data and programs used to produce the distribution clusters is essential for a full evaluation of how the model performs. However, even a cursory look at some of the clusters in GTE's wire centers casts doubt on the precision of the process. For example, in the HMTN wire center for GTE Alabama, the model contains a cluster of 14.6 square miles, 232 lines, and 182 locations. These locations are represented by lots with dimensions 1,057 x 2,115 (51 acre lots). Not only does such an outcome raise the question of how precisely customers are located

[72] *Model Documentation*, p. 28, footnote 31.